Maximum Likelihood Estimation

Estimation

Say we have data on home ownership in North Carolina, collected as yes/no

Owns a home (1=Yes)
0
1
1
…
0

We want to know about the proportion of home ownership in the state. Using the data, what might be our best guess?

Natural Estimators

Similarly, if I want to know about each of the parameters given below, what might be the best guess using a sample?

- mean μ Sample mean $\bar{x}$
- median η Sample median
- standard deviation σ sample standard deviation

These **estimators** (like $\bar{x}$ for μ) are intuitive. But sometimes it's not so simple!

Maximum Likelihood Estimation

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} F_\theta$ independent & identically distributed

probability distribution w/ parameter(s) θ

Then the **joint distribution** of $(X_1, \ldots, X_n)$ is given by
$f_\theta(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_\theta(x_i)$ product b/c X_i independent

probability density function

Now, we define the **likelihood function** as a function of the fixed population parameter $\theta : L(\theta) = f_\theta(X_1, \ldots, X_n) = \prod_{i=1}^{n} f_\theta(X_i)$
We want to find the value of θ that maximizes the likelihood function and call it $\hat{\theta}$

Illustration

Going back to our example of home ownership, let's say we have n=100, where 70 own their home. What distribution can we assume for this variable?

We can use the Bernoulli distribution, so
$\underbrace{\prod_{i=1}^{n} \underbrace{p^{x_i}(1-p)^{1-x_i}}_{\text{pmf Bernoulli}}}_{\text{joint probability distribution}}$ and $\underbrace{L(p) = p^{70}(1-p)^{30}}_{\text{Likelihood Function}}$

Illustration

We can plot $L(p) = p^{70}(1-p)^{30}$ over a range of values of p

General mechanics for finding the MLE

1. Assume a distribution for the population
2. Define the likelihood: joint probability of the data
3. Take the log of the likelihood
4. Find the value of θ that maximizes the likelihood:
    1. Differentiate the log-likelihood with respect to θ
    2. Set the derivative to 0
    3. Solve for θ

① Take the log
② Differentiate wrt p
③ set to 0 & solve for p

Example

$L(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$

$$l(p) = \sum_{i=1}^{n} x_i \log p + (1-x_i)\log(1-p)$$ ① Take the log

$$\frac{\partial l}{\partial p} \quad \frac{\sum x_i}{p} - \frac{\sum(1-x_i)}{1-p}$$ ② Differentiate wrt p

$$\frac{\sum x_i}{p} - \frac{\sum(1-x_i)}{1-p} = 0$$

$$\frac{(1-p)\sum x_i - p\sum(1-x_i)}{p(1-p)} = 0$$

③ Set to 0 and Solve for p

Find a common denominator

$$(1-p)\sum x_i - p\sum(1-x_i) = 0$$

$$\sum x_i - p\sum x_i - np + p\sum x_i = 0$$

$$\sum x_i - np = 0$$

$$\hat{p} = \frac{\sum x_i}{n}$$

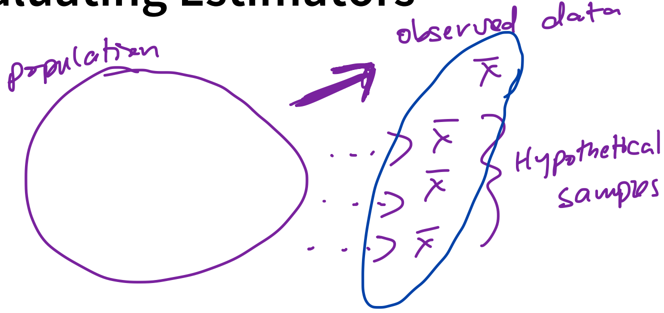
Add a "hat" ^ to denote the estimator

$$\frac{\sum_{i=1} x_i}{n} = \text{sample mean}$$

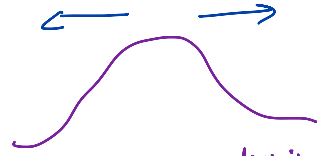
Evaluating Estimators

- Bias: Average distance from the population parameter
  - Bias($\hat{\theta}$) = $\mathbf{E}[\hat{\theta}] - \theta$
- Standard Error (SE): standard deviation of the estimator (i.e., SD of the sampling distribution)
  - STD($\hat{\theta}$)= $\sqrt{\mathbf{E}[\hat{\theta} - \mathbf{E}(\hat{\theta})]^2}$
- Mean Squared Error (MSE): combines standard error and bias
  - MSE($\hat{\theta}$)= $\mathbf{E}[\hat{\theta} - \theta]^2$ = SE^2 + Bias^2
- Consistency: the estimate converges to the true value as $n \to \infty$

Evaluating Estimators



Standard error (θ) = standard deviation of the sampling distribution



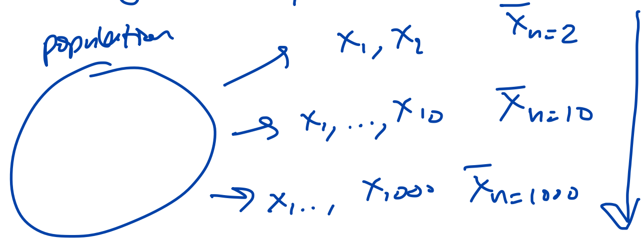
$\sigma/\sqrt{n}$ CLT

(N if CLT applies)

Bias: $E[\hat{\theta}]$

on average, am I getting the true value

Consistency



Note: Law of large numbers says the sample mean is consistent

As n increases, am I getting closer to the true value

Example

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} Poisson(\lambda)$

Derive the MLE, $\hat{\lambda}$, and determine if this is an unbiased estimator of λ

$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$

$$L(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$ ① write the likelihood

$$\ell(\lambda) = \sum_{i=1}^{n} x_i \log\lambda - \lambda - \log x_i!$$ ② Take the log

$$\frac{d\ell}{d\lambda} = \frac{\sum x_i}{\lambda} - n$$ ③ differentiate wrt λ

$$0 = \frac{\sum x_i}{\lambda} - n$$ ④ Set to 0 and solve

$$\hat{\lambda} = \frac{\sum x_i}{n}$$

$$E[\hat{\lambda}] = E\left[\frac{\sum x_i}{n}\right] = \frac{\sum E[x_i]}{n}$$

$E[x_i] = \lambda$ and $\sum_{i=1}^{n} \lambda = n\lambda$

$$= \frac{n\lambda}{n}$$

$$= \lambda \checkmark$$

unbiased